International Journal of Engineering and Computer Science Applications (IJECSA)
Vol. 4, No. 1, March 2025, pp. 13∼28
ISSN: 2828-5611
DOI: doi.org/10.30812/ijecsa.v4i1.4832

Comparison of Lexicon-based Methods and Bidirectional Encoder Representations for Transformers Models in Sentiment Analysis of Government Debt Market Movements

Firda Rachmawati, Ulil Azmi, Rahmania Azwarini
Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia

Article Info

Article history:

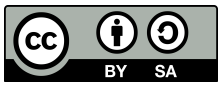
Received January 21, 2025
Revised January 23, 2025
Accepted February 04, 2025

Keywords:

Bidirectional Encoder Representations for Transformers,
Government Bonds,
Lexicon-based,
Natural Language Processing,
Sentiment Analysis.

ABSTRACT

The State Budget of Indonesia (APBN) is the main tool for implementing fiscal policies and serves as a budgeting guideline for development execution in Indonesia. One of the funding sources in budget financing is Debt Financing, which consists of Government Securities (SBN) issuance and Loans. Overall, Government Securities (SUN) contributes IDR 5,824.34 trillion, highlighting its significant proportion in debt financing. Understanding public sentiment toward SUN is essential in developing effective government policies. This research conducts sentiment analysis on tweets from the social media X over the past 7.75 years to assess public perception and propose strategic recommendations. This research aims to compare the BERT model and the Lexicon-based method to determine which achieves the highest accuracy in sentiment analysis. The findings can help the government develop strategies for issuing SUN, especially in improving public involvement and investor trust. This research method is based on deep learning pre-trained Bidirectional Encoder Representations from Transformers (BERT) model, specifically IndoBERT, with fine-tuning and a Lexicon-based approach utilizing the InSet lexicon. The results of this research are as follows: on the overall tweet dataset, the BERT model with optimal hyperparameters outperformed the Lexicon-based method, achieving an accuracy of 70.28% compared to 55.77%. Similarly, on an annual basis, BERT exhibited higher accuracy than the Lexicon-based method, except in 2021. Public sentiment on SUN in social media X is categorized as 49% positive, 30% neutral, and 21% negative. These findings indicate a generally favorable perception of SUN but highlight areas for improvement in public communication. Overall, the BERT model demonstrates superior performance over the Lexicon-based method. Considering the opportunities available, the government could leverage social media through Key Opinion Leaders and enhance transparency in explaining policies such as Tapera. This approach could maximize public participation in investing in SUN in Indonesia.



Corresponding Author:

Firda Rachmawati, +6285777246184
Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia,
Email: firdarachmawati3@gmail.com

How to Cite: F. Rachmawati, U. Azmi, and R. Azwarini, "Comparison of Lexicon-based Methods and Bidirectional Encoder Representations for Transformers Models in Sentiment Analysis of Government Debt Market Movements", *International Journal of Engineering and Computer Science Applications (IJECSA)*, vol. 4, no. 1, pp. 13-28, Mar. 2025. doi: doi.org/10.30812/ijecsa.v4i1.4832.

1. INTRODUCTION

The Government of Indonesia is important in implementing sustainable and effective fiscal policies to maintain economic stability and growth. The State Budget (APBN) is the main tool for implementing fiscal policies and serves as a budgeting guideline for the implementation of development in Indonesia. Based on the 2024 APBN structure, the 2024 state expenditure budget amounts to IDR 3,325.1 trillion, with state revenue at IDR 2,802.3 trillion and budget financing at IDR 522.8 trillion. Funding sources in budget financing are grouped into Debt Financing and Investment Financing. In 2024, the planned SBN financing is IDR 666.4 trillion, while loan financing amounts to only IDR 18.4 trillion. On average, Government Securities (SBN) were effective and contributed to covering the State Budget (APBN) deficit financing during the 2012–2021 period [1]. As of the end of June 2024, the government debt outstanding position is recorded at IDR 8,444.87 trillion, reaching 39.13% of GDP, which remains consistently below the safe limit of 60%. Based on its instruments, the composition of government debt consists of SBN amounting to 87.85% of the total government debt, with the remainder being loans. In detail, Domestic SBN contributes IDR 5,967.7 trillion, consisting of Government Securities (SUN) at IDR 4,732.71 trillion and Sharia Government Securities (SBSN) at IDR 1,234.99 trillion. SUN holds a significant proportion of debt financing, and its market movement is larger and more dynamic. This large proportion makes the SUN market an interesting subject for analysis. The market movement has the potential to be influenced by various factors experienced by the public. Public sentiment can be obtained from various sources, such as social media, enabling further analysis of SUN price movements using lexicon-based methods and machine learning.

Sentiment analysis, one of the key components of Natural Language Processing (NLP), identifies emotions based on text. In general, there are two approaches to sentiment analysis: the lexicon-based approach and the machine learning-based approach. Lexicon-based Methods can be divided into Dictionary and Corpus-based Approach [2]. The machine learning approach uses manually classified datasets as training data to automatically classify text, while the lexicon-based approach relies on an opinion dictionary (lexicon) for classification [3]. Lexicon-based was conducted by Setiawan [4] in his study titled Public Sentiment Analysis on Twitter Regarding the Astana Anyar Police Station Suicide Bombing Using the SVM Algorithm with VADER and InSet Lexicons. The results showed that sentiment analysis using the SVM algorithm with VADER labeling achieved an accuracy rate of 94%, higher than InSet at 93%. However, this study still uses InSet because the data used are tweets in the form of words or sentences in Indonesian, so it would be ineffective if it still had to be translated into English like the VADER mechanism. The Lexicon-based method has been developed for a long time and is still widely used by researchers in natural language processing. Sentiment analysis using the InSet Lexicon was conducted by Musfiroh et al. [3] on a Twitter dataset discussing online lectures, demonstrating good performance with an accuracy of 79.2%. Another study by Faizal [5] compared the Lexicon-based method with Naïve Bayes, yielding an accuracy of 65%, while the accuracy without the Lexicon-based approach was 64%. A similar study by Zilvi et al. [6] utilized the BERT model for Indonesian text classification and testing using a confusion matrix, achieving an accuracy of 85%. However, their study employed automatic labeling with the Vader Lexicon and balanced classes using Random Oversampling.

The next approach, Bidirectional Encoder Representations for Transformers (BERT), has been widely used for sentiment analysis to determine human emotional ideas. Deep learning-based sentiment analysis has significant potential in extracting valuable sentiment insights from textual data, benefiting various applications [7]. BERT-based models perform better than traditional models such as SVM, Naive Bayes, and LSTM [8]. Based on the study titled Sentiment Analysis of Customer Reviews on the Ruang Guru Application Using the BERT Method, the pre-trained BERT model is highly effective for sentiment analysis implementation [9]. The use of the BERT model in sentiment analysis of government policies was previously conducted by Kurniawan et al. [10] In their study titled Sentiment Analysis of Electronic System Operator (PSE) Policies Using the BERT Algorithm, accuracy rates were achieved at 69%, 55%, and 55%. Research utilizing the BERT model has been continuously evolving in recent years. A sentiment analysis study on social media data using the BERT model was conducted by Tabinda Kokab et al. [11], demonstrating that the CBRNN model with a BERT foundation was the most effective compared to other models. The following year, Chandradev et al. [12] conducted a similar study on review analysis using a simpler BERT model, achieving an accuracy of 91.40%. In other studies, The IndoBERT model performs well with Indonesian text, achieving an accuracy of 85% on validation training and 86% on testing training [13]. BERT was designed as a pre-trained model that is deeply bidirectionally trained on unlabeled text data by incorporating both left and right contextual layers, allowing it to be adapted to machine learning tasks with the addition of just a single layer. Its rapid development has enabled the model to be applied to sentiment analysis in the Indonesian language, yielding high-accuracy results [14].

Previous research has not resolved some gaps, namely the limited studies on sentiment analysis using BERT for government policies, particularly related to Government Securities (SUN). The role of SUN is very important for the sustainability of the Indonesian economy. Increasing Government Securities (SBN) will increase economic growth [15]. Additionally, comparisons between traditional classification methods, such as Lexicon-based approaches, and modern machine learning models, like BERT, in the context of government policies remain underexplored. The difference between this research and the previous one is that this research directly compares these two approaches to determine which method offers more accurate sentiment analysis. By obtaining the accu-

rate results of both methods, this research identifies the more effective model, which can be further optimized and used for strategic decision-making in SUN issuance. This research contributes to the scientific field by expanding sentiment analysis research on government financial instruments, particularly in Indonesia, where such studies are still limited. The benefits of this research include providing insights into sentiment analysis using BERT and Lexicon-based methods, specifically concerning SUN movements as a key government policy, and offering a summary of public sentiment on SUN-related policies, which dominate deficit financing in Indonesia. These insights can serve as a valuable reference for policymakers in formulating future issuance strategies.

The government can utilize this technological advancement to evaluate policies and develop budget optimization strategies. The growth of SBN will influence Indonesia's economic growth [16] and Government Securities (SUN) positively impact Indonesia's economic growth both in the short and long term. SUN can be sold through the capital market mechanism, and basically, a person participates in the implementation of a SUN auction for the sole purpose of investment [17]. Public sentiment, views, and trust in purchasing SUN are important to analyze, especially since SUN in SBN dominates debt financing for budget deficits. This research uses sample tweet data from the social media platform X or Twitter over the past 7.75 years. Samples of public sentiment were obtained from social media X or Twitter using keywords such as "Surat Utang Negara," "SUN," "Obligasi," "ORI," "Obligasi Ritel," "SBR," and "Saving Bond." The methods used are the pre-trained deep learning model Bidirectional Encoder Representations from Transformers (BERT), namely IndoBERT, with the fine-tuning method and Lexicon-based using the InSet lexicon. The researchers hope that the results of this research can be utilized by relevant parties and the general public as policy evaluation material, strategic considerations to optimize SUN as the dominant debt instrument in financing budget deficits in Indonesia, and as a study for academics who require it.

2. RESEARCH METHOD

This research uses data scraping from the social media platform X, with the dependent variable being the sentiment classification of tweets categorized as positive, negative, or neutral. The independent variable, consisting of tweet texts related to SUN, serves as the primary input in the sentiment analysis of this research. The flow of this research is illustrated in Figure 1. The research begins with a literature review on the Lexicon-based method using InSet and BERT for sentiment analysis in the Indonesian language. The collected data is ensured to be relevant to the topic and sufficiently extensive for training the BERT model. Subsequently, the data undergoes manual labeling to provide accurate ground truth data. Positive labels are assigned to tweets containing positive testimonials about the purchase and redemption of SUN, explanations or education, calls to purchase SUN, information on issuance that highlights its unique features, and services from distribution partners that assist consumers. Negative labels are assigned to tweets containing mockery, questions indicating a lack of literacy or information, and dissatisfaction with existing systems and policies. Meanwhile, neutral labels are assigned to tweets with incomplete information, announcements lacking significant details, and general explanations about SUN. The next stage is pre-processing, which begins with cleansing to remove unnecessary characters such as punctuation marks, numbers, and symbols. The second step is case folding, which converts all text to lowercase for consistency. The third step is stemming, which reduces words to their root forms to simplify the analysis. Lastly, tokenizing is performed to split the text into words or tokens to facilitate further analysis.

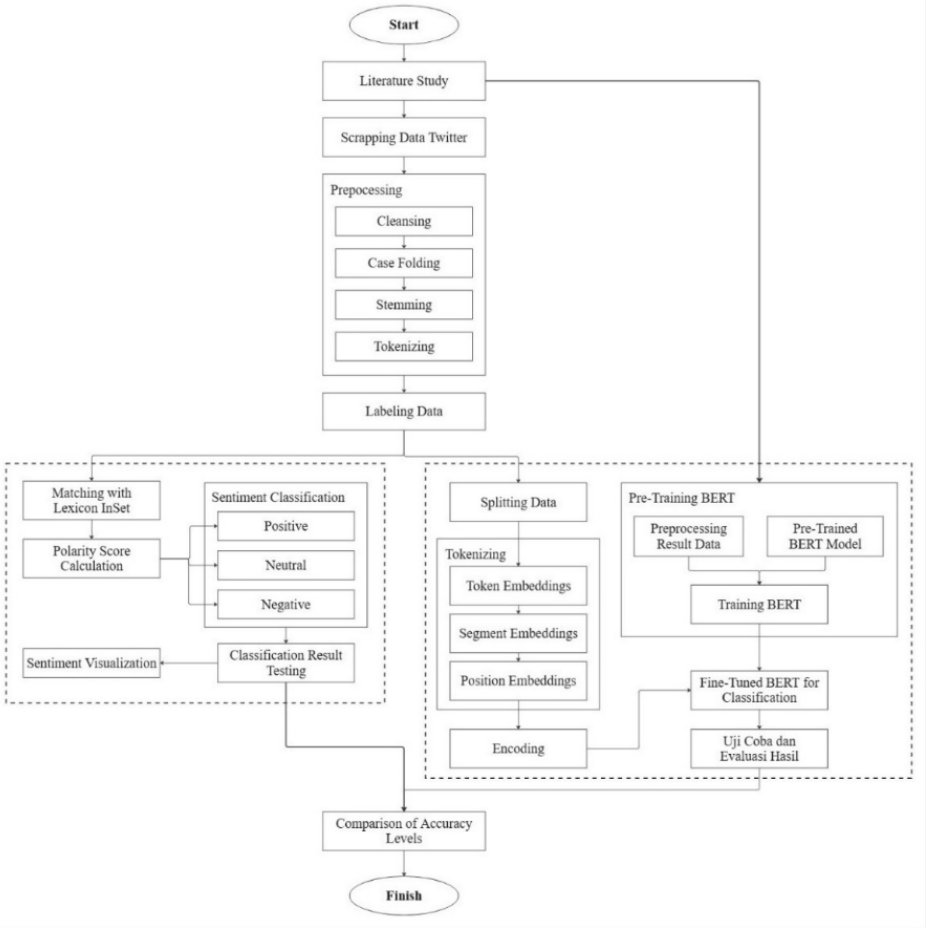


Figure 1. Research Flow

Sentiment analysis is conducted using two methods. The first method is Lexicon-based, where positive and negative word lists from the InSet lexicon are matched with input tweets to calculate the polarity score. The sentiment classification results are visualized and evaluated by comparing them with the labeled dataset to determine the accuracy and effectiveness of the method used. The researchers utilized the Indonesian Sentiment (InSet) Lexicon, which contains 3,609 positive words and 6,609 negative words in Indonesian, each assigned a polarity score ranging from -5 to +5 to classify sentiment types. The calculation of the polarity score is performed by summing up the weights of all detected words, which are then classified into sentiment types through an algorithm: classified as positive if the polarity score is greater than 0, negative if the polarity score is less than 0, and neutral if the polarity score is exactly 0. The second method is BERT, which begins with splitting the data into training, validation, and testing sets. Weight calculations using a weighted loss function are performed to enhance the model's sensitivity to minority classes with Equation 1.

$$class\ weight = 1 - \frac{number\ of\ data\ per\ class}{total\ data\ in\ the\ dataset} \tag{1}$$

Tokenization and pre-training of BERT proceed through embedding stages, followed by model training using IndoBERT for the sentiment analysis classification task. During the embedding stage, the first token of each sequence is the Special Class Classification token ([CLS]) to distinguish each sentence, and it ends with a Separate token ([SEP]) to signify the end of the sentence. Token embeddings are used to represent the meaning of words in a sentence. Segment embeddings are added to each token to differentiate words from one sentence to another, and Position embeddings are added to indicate the position of tokens in a sentence. The sum of Token Embeddings, Segment Embeddings, and Position Embeddings is used as input to the Encoder mechanism [5], represented as a matrix X in Equation 2. The elements of the matrix X are shown in Equation 3.

$$X = TE + PE + SE \tag{2}$$

$$X = \begin{bmatrix} x_{11,} & \cdots & x_{1N,} \\ x_{21,} & \cdots & x_{2N,} \\ \vdots & \vdots & \vdots \\ x_{N1,} & \cdots & x_{NN,} \end{bmatrix} \tag{3}$$

The matrices Q, K, and V are matrices that contain all the queries, keys, and values, and $X = (x_1, x_2, \ldots, x_N)$ represents the sequence of tokens x_i for $i = (1,\ 2,\ \ldots,\ N)$. W^Q, W^K, and W^V are weight matrices. Each matrix vector is defined as in Equation 4. The similarity score, representing the relationship between tokens x_i and x_j, is computed using scaled dot-product attention. The similarity score S_{ij} is calculated by the following Equation 5. Here, d_k is the dimension of the key vector, and $\forall_j = 1,\ \ldots,\ N$. The similarity scores are then transformed into probabilities using the softmax function. The softmax function is applied to the calculated similarity scores to obtain the attention weights. For the calculation, $\forall_k = 1,\ \ldots,\ N$. The attention weight of the token i for token j is calculated by the following Equation 6.

$$\begin{aligned} q_i &= x_i\, W^Q \\ k_i &= x_i\, W^K \\ v_i &= x_i\, W^V \end{aligned} \tag{4}$$

$$S_{ij} = \frac{q_i\ .\ k_i^T}{\sqrt{d_k}} \tag{5}$$

$$a_{ij} = softmax\,(s_{ij}) = \frac{exp\,(s_{ij})}{\sum_{k=1}^{N} exp\,(s_{ik})} \tag{6}$$

The attention weight a_{ij} is used to calculate the output for each token. The following equation describes how the output o_i is generated by summing the attention weights a_{ij}, each multiplied by the value matrix V and the previous token x_j using the value weight matrix W^V. The calculation for the output of the token i is performed by the following Equation 7. The self-attention mechanism as a whole can be represented by the following Equation 8. This mechanism allows the model to generate contextual representations that capture the relationships between tokens in the input sequence, as outlined in the previous steps. The second mechanism is multi-head attention, which enhances the flexibility of the self-attention process. The mechanism for each head denoted as $head_i$, is defined by Equation 9.

$$o_i = \sum_{j}^{N} a_{ij}\ v_j \tag{7}$$

$$Attention\,(Q,\ K,\ V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right) V \tag{8}$$

$$head_i = attention\left(QW_i^Q,\ KW_i^K,\ VW_i^V\right) \tag{9}$$

$$H' = MultiHead(Q, K,\ V) = Concat\,(head_1,\ head_2, \ldots,\ head_h)\ W^O \tag{10}$$

The query, key, and value vectors for each attention process are computed with a linear projection of the original input using the projection matrices $W_i^Q,\ W_i^K,\ W_i^V\ \in \mathbb{R}^{2d_h \times d_k}$, where $d_k = 2d_h\ /h$. The Concat operation in Equation 2 is MultiHead, which combines the results of all heads, each of which has a dimension of d_k. This mechanism allows the model to capture various interactions between tokens in the context. The model is then refined through fine-tuning with a combination of hyperparameters to achieve optimal results. The model is tested using a test dataset to predict sentiment categories. The hyperparameter combination includes a batch size of 16 and 32, epochs of 2, 3, and 4, and learning rates of 2e-0,5,3e-0,5, and 5e-0,5.

The next stage after multi-head attention is processing each position separately and identically in the sequence through a fully connected feed-forward network, which is contained in each layer of both the encoder and the decoder. This mechanism uses ReLU activation as in Equation 11. In this equation, the input vector is represented by x, W_1 and W_2 are the weights of the first and second linear transformations, and b_1 and b_2 are the bias vectors for the first and second linear transformations. Next, the intent label y^{intent} is predicted based on the semantic representation of the sentence H' as in Equation 12. The weight and bias matrices in the intent classifier layer are represented by W^{int} and b^{int}. The softmax function is used to convert the model output into a relative probability distribution across the possible classes, allowing for the selection of the class with the highest probability as the final prediction.

$$FFN\,(x) = \max\,(0, x\,W_1 + b_1)W_2 + b_2 \tag{11}$$

$$y^{intent} = softmax\,\left(W^{int}\,H' + \,b^{int}\right) \tag{12}$$

The model evaluation uses a confusion matrix and performance analysis with class adjustment as Tabel 1. The performance evaluation is conducted by comparing both methods in terms of accuracy and efficiency to determine the best approach for sentiment analysis related to the movement of SUN in Indonesia. This research uses three sentiment categories, so the confusion matrix is adjusted to become a multiclass confusion matrix. The confusion matrix consists of four types of performance evaluation, namely Accuracy, Precision, Recall, and F1 Score, each of which has a formula such as Equation 13-16.

Table 1. Confusion Matrix

Evaluasi	Confusion Matrix	Multiclass Confusion Matrix	
Accuracy	$\frac{TP+TN}{FP+FN+TP+TN}$	$\frac{TPos+TNet+TNeg}{TPos+TNet+TNeg+FPos+FNet+FNeg}$	(13)
Precision	$\frac{TP}{FP+TP}$	$\frac{TPos}{FPos+TPos}$	(14)
Recall	$\frac{TP}{TP+FN}$	$\frac{TP}{TPos+FNet+FNeg}$	(15)
F1 Score	$2 \times \frac{Presisi \times Recall}{Presisi + Recall}$	$2 \times \frac{Presisi \times Recall}{Presisi + Recall}$	(16)

3. RESULT AND ANALYSIS

The findings of this research are that the BERT model (IndoBERT) demonstrated better performance in sentiment classification across the entire tweet dataset, achieving an accuracy of 70.28%. In comparison, the lexicon-based InSet method attained only 55.77%. The BERT model consistently outperformed the lexicon-based method across most datasets. However, both approaches proved fairly effective in classifying tweets into three sentiment categories. The results of this research are in line with or supported by previous studies. Sentiment analysis using deep learning-based techniques holds great promise in unlocking valuable sentiment insights from textual data [7]. The BERT model's accuracy of 70.28% aligns with the findings of Kurniawan et al. [10], who conducted sentiment analysis on government policy (PSE) using BERT and reported an accuracy range of 55–69%. Meanwhile, the lexicon-based method's accuracy of 55.77% indicates that the InSet lexicon is fairly effective in correctly classifying more than half of the dataset into the appropriate sentiment categories. This finding is consistent with Faisal [5], who demonstrated that sentiment analysis using a lexicon-based approach performs better than not using it at all. Compared to previous research, this study is unique in that it directly compares the lexicon-based method and the BERT model in the context of SUN. While prior studies have explored sentiment analysis using either lexicon-based methods or BERT models individually, none have examined both approaches in this specific domain. Additionally, previous research has focused on either sentiment analysis methods or the role of SUN in public finance, but not their intersection. Therefore, this study fills an important gap by integrating these aspects and can serve as a foundation for future research, particularly in determining the most effective sentiment analysis method for SUN-related topics. The Twitter Search Scraper facilitates the data collection process on Apify.com and the Twitter crawler using Tweet-Harvest. Data collection is conducted by applying filters to limit the data to Indonesian tweets, using several keywords, and within the period of January 1, 2017, to September 30, 2024. Table 2 The raw data from the scraping results below shows tweet data with several specified keywords within that period.

Table 2. Raw Data Sample Tweet Data

No	Tweet
1	15. Saving bond ritel (SBR) mencapai 12.969 investor Obligasi Negara Ritel (ORI) mencapai 225.814 investor #SadarAPBN #PembiayaanSehat
2	BARU NGERTI KENAPA KALO BELI BONDS/STOCKS ITU KITA SAVING AND NOT INVESTING I AM SHOOK
3	So skrg ni kena start saving dulu..unless company nak support..tp mcm x minat nk kena bond je dgn company..huhu
4	10 temanKeu dapat berinvestasi membeli SBN berupa Obligasi Negara Ritel Indonesia (ORI), Sukuk Ritel, Savings Bond Ritel (SBR)
5	Utang Negara Rp4.274 Triliun, Tiap Orang Indonesia Tanggung Rp16 Juta http://okz.me/CYd0y
...	...
2490	Yield SBN 10 tahun Indonesia tetap kompetitif di angka 6,72%, menunjukkan stabilitas pasar obligasi yang terus didorong oleh kebijakan fiskal yang responsif.
2491	”Investasi cerdas dengan Obligasi Negara Ritel ORI026! Pilihan tepat untuk masa depan finansial yang lebih efisien. Pemerintah kembali meluncurkan investasi ORI026 dengan instrument investasi yang aman, dan dijamin negara. Dapatkan kupon bersifat tetap (fixed rate) hingga 6,40% per tahun sehingga tidak terpengaruh oleh fluktuasi suku bunga pasar. Pilih tenor sesuai dengan kebutuhan finansialmu: • ORI026-T3: Tenor 3 tahun dengan besar kupon 6,30% per tahun • ORI026-T6: Tenor 6 tahun dengan besar kupon 6,40% per tahun per tahun Yuk, segera lakukan pembelian untuk nikmati cashback hingga 0,20% maks.Rp27,5 juta, berlaku hingga 10 Oktober 2024. Agar lebih praktis dan mudah berinvestasi, beli ORI026 secara online via M2U ID App atau kunjungi kantor cabang Maybank terdekat. Beli sekarang dan jangan sampai terlewat batas penawarannya! Periode penawaran 30 September 2024 – 24 Oktober 2024 Info lebih lanjut : maybank.co.id/InvestORI026 Info panduan pemesanan ORI026 via M2U ID App : maybank.co.id/ORI026viaM2U Syarat dan ketentuan berlaku #MyBank #InvestasiNegeriku #PilihanBerharga #DemiMasaDepanBangsa”
2492	Anjir gw selalu merasa pas covid invest saham tuh bener2 ga berasa untungnya banget, aplg dibanding surat utang negara. Ternyata bener aja.
2493	Obligasi RI bakal terimbas positif dari pemangkasan suku bunga Fed dlvr.it/TDrBFx
2494	Ga cuma soal rate BI, bank konvensional lbh banyak salurkan kredit ke instrumen pemerintah, SUN, SUKUK, obligasi pemerintah, pendanaan proyek pemerintah, kredit rakyat zonk.

The data from each keyword search is then saved in Excel format and consolidated into a single dataset. The researcher creates two datasets from the obtained data: the first dataset consists of tweets for each period/year, and the second dataset includes all tweets within the specified time range, effectively combining the first dataset. Data unrelated to the topic of Government Securities (SUN) are manually removed from the dataset by the researcher. In total, 2,494 tweets are prepared for processing, and the distribution of the data is shown in Figure 2.

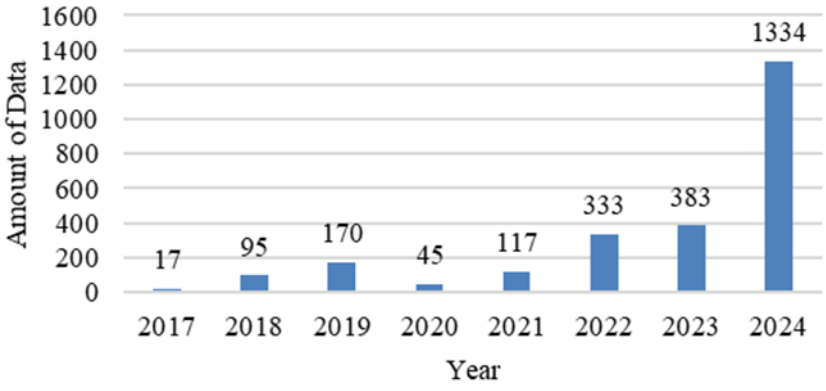


Figure 2. Distribution of Tweet Data by Year

3.1. Data Labeling (Ground Truth)

Next, the tweets in each dataset are manually labeled. Table 3 below shows some examples of tweets labeled positive, neutral, and negative. The labels are created by the researcher not only in the form of the words "Positive," "Neutral," and "Negative" but also indicated by the numbers 1, 0, and -1 to facilitate the subsequent sentiment analysis process with Lexicon and BERT methods. Table 4 shows the labeling results for the entire tweet dataset and for each year.

Table 3. Manual Labeling (Ground Truth)

No	Tweet	Sentiment
...	...	...
2313	Surat berharga negara/obligasi/surat utang, kayak minjemin uang ke negara nanti kita dapet bunga dari pinjaman itu.. Obligasi/SBN salah satu investasi teraman dan pajaknya lebih kecil dari deposito. SBN 10%, deposito 20%.	Positive
...	...	...
2324	Surat Utang Negara (SUN) masih diburu oleh investor dari dalam maupun luar negeri. dlvr.it/TCnLfR	Positive
...	...	...
1550	bentuk obligasi bermacam bisa surat berharga negara surat utang negara dll yang berhubungan sm pemerintah.	Neutral
...	...	...
1951	Jadi bagaimana pengaruh el nino kepada obligasi ritel apakah akan naik atau turun	Neutral
...	...	...
1787	@liputan6dotcom Dikira rakyat bodoh semua aoa ya? Duit tapera dipake utk beli SUN(surat utang negara) nah duit yg masuk dr SUN dipake utk biayai belanja pemerintah bs jd dibuat utk bangun IKN buat makan siang gratis dsb. Duit tapera utk dipinjam pmrnth	Negative
...	...	...
2340	Embeerrr! Bisa-bisanya jualan surat utang melulu. Kek sebulan sekali ada aja surat utang yg dikeluarin. Idup udah dirugiin negara, rakyat nya jg disuruh minjemin duit ke negara. Fak lah bgst	Negative
...	...	...

Table 4. Proportion of Sentiment in Ground Truth Data for the Entire Tweet Dataset and by Year

Year	Positive	Neutral	Negative
2017	7	7	3
2018	50	36	9
2019	70	87	13
2020	19	19	7
2021	54	43	20
2022	177	110	46
2023	207	134	42
2024	650	311	373
All	1234	747	513

3.2. Pre-Processing

The data obtained from Twitter scraping is raw data, so it requires further processing, namely data preprocessing, to eliminate unnecessary elements and produce a clean and relevant dataset. This process is carried out for each tweet in every dataset. Table 5 below shows an example of data preprocessing on one of the tweets.

Table 5. Data Preprocessing Stage

Original Tweet	Preprocessing Result	Case Folding Result	Stemming Result	Tokenizing Result
@DeniSetiabudi15 Halo Bpk Deni apabila yg Bpk maksudkan adalah pemesanan Saving Bond Ritel Seri SBR-004 dpt kami informasikan SBR-004 hanya dapat dipesan melalui Layanan Mandiri SBN Ritel Online yaitu portal pemesanan SBN https://t.co/jimkFkioxg. (1)	Halo Bpk Deni apabila Bpk maksudkan adalah pemesanan Saving Bond Ritel Seri SBR dpt kami informasikan SBR hanya dapat dipesan melalui Layanan Mandiri SBN Ritel Online yaitu portal pemesanan SBN	halo bpk deni apabila bpk maksudkan adalah pemesanan saving bond ritel seri sbr dpt kami informasikan sbr hanya dapat dipesan melalui layanan mandiri sbn ritel online yaitu portal pemesanan sbn	halo bpk den apabila bpk maksud adalah mesan saving bond ritel seri sbr dpt kami informasi sbr hanya dapat pes lalu layan mandiri sbn ritel online yaitu portal mesan sbn	['halo', 'bpk', 'den', 'apabila', 'bpk', 'maksud', 'adalah', 'mesan', 'saving', 'bond', 'ritel', 'seri', 'sbr', 'dpt', 'kami', 'informasi', 'sbr', 'hanya', 'dapat', 'pes', 'lalu', 'layan', 'mandiri', 'sbn', 'ritel', 'online', 'yaitu', 'portal', 'mesan', 'sbn']

3.3. Lexicon-based Methods

Automatic labeling using the InSet lexicon was performed on both the entire dataset and yearly datasets. Table 6 is an example of the calculation using the lexicon. The total score in the above example is 7, indicating that the data falls into the positive sentiment category. The total score in the above example is 7, indicating that the data falls into the positive sentiment category. This process is also applied to the entire tweet dataset as well as to each yearly dataset. Table 7 below presents the labeling results for the entire dataset.

Table 6. Example of Lexicon-based InSet Labeling Calculation

Preprocessing Tekonizing	Word Count	Word Category	Weight
'halo', 'bpk', 'den', 'apabila', 'bpk', 'maksud', 'adalah', 'mesan', 'saving', 'bond', 'ritel', 'seri', 'sbr', 'dpt', 'kami', 'informasi', 'sbr', 'hanya', 'dapat', 'pes', 'lalu', 'layan', 'mandiri', 'sbn', 'ritel', 'online', 'yaitu', 'portal', 'mesan', 'sbn'	halo	Positif	1
	maksud	Positif	3
	informasi	Positif	2
	hanya	Negatif	-3
	dapat	Positif	2
	lalu	Positif	3
	layan	Positif	2
	mandiri	Negatif	-3

Table 7. Example of Lexicon-based InSet Labeling Calculation

Year	Positive	Neutral	Negative
2017	9	2	6
2018	67	7	21
2019	102	15	53
2020	28	7	10
2021	86	2	29
2022	238	21	74
2023	265	33	85
2024	762	64	508
All	1557	151	786

The accuracy of the classification results is tested by comparing the classification results from the Lexicon-based method with InSet to the actual classification results. This testing using a confusion matrix, with the results as in Figure 3. A confusion matrix is also performed for the tweet dataset for each year. Based on the results, performance evaluation is obtained, which includes accuracy, precision, sensitivity, and F1-score. Table 8 is a summary of accuracy level for Lexicon Based InSet. Based on the results, overall, the model correctly predicted 55.77% of the time. The performance of lexicon-based InSet model on the yearly dataset tends to be stable at 40-50%, with an increasing accuracy rate each year, except for a slight decrease in 2024. This indicates that lexicon-based approach is solely based on the InSet lexicon and does not depend on the amount of data.

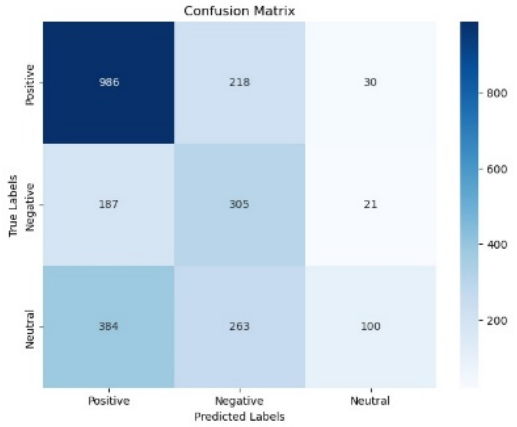


Figure 3. Confusion Matrix Lexicon Based InSet Entire Tweet Dataset

Table 8. Accuracy Level of Lexicon Based InSet for the Entire Dataset

Year	Amount of Data	Accuracy Level
2017	17	0,41176
2018	95	0,55789
2019	170	0,46471
2020	45	0,48889
2021	117	0,54701
2022	333	0,57057
2023	383	0,62141
2024	1334	0,55322
All	2494	0,55774

3.4. BERT Models

The Transformer is a deep learning model with an encoder-decoder structure for translation tasks. BERT, a multi-layer bidirectional Transformer encoder, maps words to word vectors based on context. Unlike earlier models, BERT is pre-trained to capture deep bidirectional language representations across all layers [18]. Sentiment analysis using the BERT model begins with splitting the data into training, validation, and testing datasets. The training and testing data proportion of 70:30 yielded the best results among several other cases [9]. The division of training data into validation and test data allows for proportions such as 70:15:15 or 70:20:10. Based on these schemes, the researchers implemented the model and fine-tuned it using the best hyperparameter combinations according to the theory for each scheme, identifying the configuration that produced the highest accuracy. As a result, this research divided the data into 70% for training (1,745 tweets), 20% for validation (500 tweets), and 10% for testing (249 tweets). The data split revealed significant class imbalance, necessitating the use of a weighted loss function for each class to assign higher weights to misclassified minority class errors. Table 9 is shown the weights assigned to each class for every dataset.

Table 9. Class Weights for Each Class in Each Tweet Dataset

Year	Class Weight		
	Neutral	Positive	Negative
2017	0,5882	0,5882	0,8235
2018	0,6211	0,4737	0,9053
2019	0,4882	0,5882	0,9235
2020	0,5778	0,5778	0,8444
2021	0,6325	0,5385	0,8291
2022	0,6697	0,4685	0,8619
2023	0,6501	0,4595	0,8903
2024	0,7669	0,5127	0,7204
All	0,7005	0,5052	0,7943

The tokenization process converts the obtained tokens into numerical indices according to the predefined vocabulary by adding special tokens [CLS] and [SEP]. Padding with the [PAD] token and attention mask is also applied to maintain consistent input length and distinguish between the original tokens and padding. Table 10 shows each stage in this process. The sentiment analysis classification model used is IndoBERT Base developed by IndoLEM, specifically ”indolem/indobertweet-base-uncased.” This model is effective, faster, and lighter. The tokenization process results are used as input embeddings in the IndoBERT input layer. This process consists of Token Embedding (TE), Positional Embedding (PE), and Segment Embedding (SE). This is an example of the embedding process for the shortest tweet, ”sbn pak.”

Table 10. Tokenization Process on a Sample Tweet

Process	Result
Text	halo bpk den . . . portal mesan sbn
Token	[’halo’, ’bpk’, ’den’, . . . ’portal’, ’mes’, ’##an’, ’sbn’]
Token	[[CLS], ’halo’, ’bpk’, ’den’, . . ., ’portal’, ’mes’, ’##an’, ’sbn’, [SEP], [PAD], . . . [PAD], [PAD]]
Token ID	[3, 18366, 8938, . . ., 2009, 1476, 13527, 4, 0, . . ., 0, 0]
Attention Mask	[1, 1, 1, 1, . . ., 1, 1, 1, 1, 0, . . ., 0, 0]

$$T = [3,\ 13527,\ 2458,\ 4]$$

$$TE = \begin{bmatrix} 0,0215 & -0,0118 & -0,0051 \\ 0,0141 & 0,0203 & 0,0704 \\ -0,0326 & -0,0541 & 0,0316 \\ 0,0111 & -0,0193 & -0,0238 \end{bmatrix},\ PE = \begin{bmatrix} 0,0545 & -0,0120 & -0,0477 \\ 0,0107 & 0,0198 & 0,0590 \\ -0,0235 & -0,0045 & 0,0494 \\ -0,0069 & -0,0426 & 0,2506 \end{bmatrix},\ SE = \begin{bmatrix} 0,0138 & 0,0123 & 0,0412 \\ 0,0138 & 0,0123 & 0,0412 \\ 0,0138 & 0,0123 & 0,0412 \\ 0,0138 & 0,0123 & 0,0412 \end{bmatrix}$$

$$X = \begin{bmatrix} 0,0898 & -0,0115 & -0,0116 \\ 0,0387 & 0,0524 & 0,1705 \\ -0,0423 & -0,0464 & 0,1221 \\ 0,0179 & -0,0496 & 0,2679 \end{bmatrix}$$

$$X' = \begin{bmatrix} 1,4434 & -0,7209 & -0,0000 \\ -0,8366 & -0,6004 & 1,4369 \\ -0,6951 & -0,7480 & 1,4431 \\ -0,4543 & -0,9593 & 1,4136 \end{bmatrix}$$

The matrix X, which is the summation of Token Embeddings (TE), Position Embeddings (PE), and Segment Embeddings (SE), needs to undergo standardization to prevent divergence or slow convergence during the model training process. A diverging model can lead to suboptimal classification results, causing tweets to fall into multiple sentiment categories simultaneously. Conversely, an overly converged model might assign tweets to a single sentiment category, with the probabilities for other categories being near zero. To address these issues, the matrix is subjected to dropout to prevent overfitting and enable the model to generalize better globally. The standardized and regularized matrix, after applying dropout, is shown by the matrix X' above. The encoder layer stage generates the matrices Q, K, and V using the matrix W, which is a 3x3 identity matrix. The Query matrix (Q), Key matrix (K), and Value matrix (V) $\in R^{l\times 3}$ represent each word in the sentence at the encoder layer.

$$Q = XW^Q = \begin{bmatrix} 1,4434 & -0,7209 & -0,0000 \\ -0,8366 & -0,6004 & 1,4369 \\ -0,6951 & -0,7480 & 1,4431 \\ -0,4543 & -0,9593 & 1,4136 \end{bmatrix},\ K = XW^K = \begin{bmatrix} 1,4434 & -0,7209 & -0,0000 \\ -0,8366 & -0,6004 & 1,4369 \\ -0,6951 & -0,7480 & 1,4431 \\ -0,4543 & -0,9593 & 1,4136 \end{bmatrix},\ V = XW^V = \begin{bmatrix} 1,4434 & -0,7209 & -0,0000 \\ -0,8366 & -0,6004 & 1,4369 \\ -0,6951 & -0,7480 & 1,4431 \\ -0,4543 & -0,9593 & 1,4136 \end{bmatrix}$$

The next stage involves calculating the similarity score S_{ij} between Q and K. This result is used to obtain the attention weight a_{ij}, which represents the importance of each token or the relevance of a specific word to other words within the context of the sentence. The calculation of similarity scores and attention weights is performed for each query, and the results are as follows:

$$S_{ij} = \begin{bmatrix} 1,5029 & -0,4473 & -0,2679 & 0,0207 \\ -0,4473 & 1,8043 & 1,7922 & 1,7247 \\ -0,2679 & 1,7922 & 1,8043 & 1,7744 \\ 0,0207 & 1,7247 & 1,7744 & 1,8042 \end{bmatrix},\ a_{ij} = \begin{bmatrix} 0,6495 & 0,0924 & 0,1105 & 0,1475 \\ 0,0349 & 0,3315 & 0,3275 & 0,3061 \\ 0,0408 & 0,3203 & 0,3242 & 0,3146 \\ 0,0549 & 0,3016 & 0,3170 & 0,3266 \end{bmatrix}$$

The next mechanism in this stage is the attention value A_i, which involves the matrix V from the previous embedding stage. The attention weight of each word obtained is multiplied by the elements of the Value matrix v_k. Using Equation 7, the calculation of the attention value for the first query A_1 as follows:

$$\begin{aligned} A_1 &= \textstyle\sum_k a_{1k}\ v_k \\ A_1 &= 0,6495 \begin{bmatrix} 1,4434 & -0,7209 & -0,0000 \end{bmatrix} + 0,0924 \begin{bmatrix} -0,8366 & -0,6004 & 1,4369 \end{bmatrix} + 0,1105 \begin{bmatrix} -0,6951 & -0,7480 & 1,4431 \end{bmatrix} + \\ &\quad + 0,1475 \begin{bmatrix} -0,4543 & -0,9593 & 1,4136 \end{bmatrix} \\ A_1 &= \begin{bmatrix} 0,7164 & -0,7479 & 0,5008 \end{bmatrix} \end{aligned}$$

In the sentiment analysis using IndoBERT, a BERT pooler is utilized at the end of each encoder layer. In the BERT pooler for classification tasks, only the embedding of the [CLS] token in A_1 is used to represent the overall information from the input as a vector for classification. The linear layer transforms the embedding of the [CLS] token using a weight matrix W, represented as an identity matrix, and a bias vector b, represented as a zero matrix.

$$
\begin{aligned}
z &= A_1 \times W + b \\
z &= \begin{bmatrix} 0,7164 & -0,7479 & 0,5008 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 \end{bmatrix} \\
z &= \begin{bmatrix} 0,7164 & -0,7479 & 0,5008 \end{bmatrix}
\end{aligned}
$$

The calculation of the BertPooler is performed by applying the tanh activation function to the obtained z. Its output is then used as input for the classifier layer, resulting in the predicted intent label by utilizing W^{int} (classifier weights) and b^{int} (bias), which can be assumed as below. The obtained results are transformed using the softmax function to indicate the class probability.

$$H' = BertPooler\ (A) = \tanh(z) = \begin{bmatrix} 0,6147 & -0,6339 & 0,4627 \end{bmatrix}$$

$$
\begin{aligned}
z &= H' W^{int} + b^{int} \\
z &= \begin{bmatrix} 0,6147 & -0,6339 & 0,4627 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} + \begin{bmatrix} -0,5 & 1 & -0,5 \end{bmatrix} \\
z &= \begin{bmatrix} 0,1147 & 0,3661 & -0,0373 \end{bmatrix}
\end{aligned}
$$

$$y^{intent} = \begin{bmatrix} 0,3180 & 0,4089 & 0,2731 \end{bmatrix}$$

Based on these results, it can be observed that the second class shows the highest probability. Thus, it is concluded that the model predicts the intent label falls into the second class, which corresponds to label 1 or positive sentiment classification. The sentiment analysis model created for the specific task requires fine-tuning to adapt the pre-trained model to better suit the classification task. Based on experimental results from various hyperparameter combinations, it was found that the combination of a batch size of 32, 4 epochs, and a learning rate of 2e-05 provided the best results, achieving the highest accuracy compared to other combinations. This combination showed a consistent decrease in training loss with increasing epochs, along with a reduction in validation loss until the end of training, indicating that the model did not experience overfitting. The fine-tuned model was subsequently tested on the testing dataset. The model used was optimized through hyperparameter tuning and dropout adjustments, resulting in optimal accuracy and a well-behaved loss curve. The test results and sentiment classification for the entire dataset is shown by Table 11.

Table 11. Proportion of Sentiment Categories from IndoBERT Model Classification on the Entire Tweet Dataset and Each Year

Year	Positive	Neutral	Negative
2017	8	3	6
2018	36	59	0
2019	84	86	0
2020	11	34	0
2021	70	47	0
2022	218	106	9
2023	220	163	0
2024	654	299	381
All	1221	763	510

Similar to the lexicon-based method, the accuracy of classification results was tested using a confusion matrix. The confusion matrix shows the accuracy of the classification results in the category label 0, which is neutral; label 1, which is positive; and label 2, which is negative. The results for the entire tweet dataset are as Figure 4.

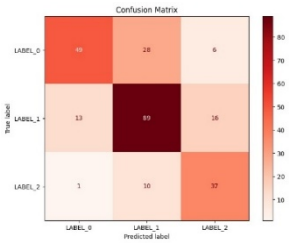


Figure 4. Confusion Matrix of IndoBERT Model on the Entire Tweet Dataset

Table 12. Proportion of Sentiment Categories from IndoBERT Model Classification on the Entire Tweet Dataset and Each Year

Year	Positive	Neutral	Negative
2017	8	3	6
2018	36	59	0
2019	84	86	0
2020	11	34	0
2021	70	47	0
2022	218	106	9
2023	220	163	0
2024	654	299	381
All	1221	763	510

3.5. Performance Evaluation Comparison

Overall, both the lexicon-based method using InSet and the IndoBERT model are fairly effective in classifying tweet data into three sentiment categories. On the entire tweet dataset, the BERT model demonstrated better performance with an accuracy of 70.28% compared to the lexicon-based InSet method, which achieved only 55.77%. The superior performance of the BERT model can be attributed to its use of deep, bidirectional contextual representations and its ability to learn from large datasets. In contrast, the lexicon-based method is limited by the predefined dictionary it relies upon, which restricts its ability to capture the nuances and complexities of sentiment in the text.

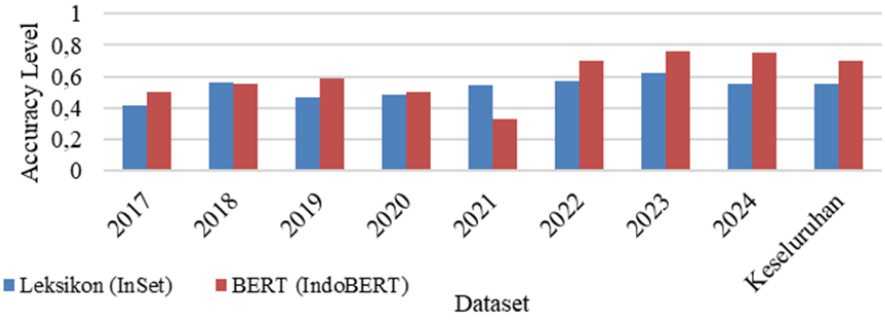


Figure 5. Comparison of Accuracy Levels Between the Lexicon-based Method and IndoBERT

Based on Figure 5, the BERT model (IndoBERT), shows higher accuracy than the lexicon-based method with InSet on most datasets. The lexicon-based method tends to remain stable, with the highest value on the 2023 dataset at 0.62% and the lowest on the 2017 dataset at 41.17%. The lexicon-based method is not significantly affected by the amount of data in the dataset because it relies on a dictionary-based rule that does not require training. The BERT model shows a performance trend that improves over time, peaking at 76.31% on the 2023 dataset, with a drastic drop to 33.33% on the 2021 dataset. This indicates that data availability influences the model's ability to perform the task optimally. The lexicon-based with InSet is suitable for smaller or medium-sized datasets, while the BERT model is better for large and complex datasets. The lexicon-based provides more consistent results, yielding stable accuracy across different dataset sizes and requiring fewer computational resources because it does not involve model training. Meanwhile, the BERT model shows performance variation, excels with complex datasets, and requires higher computational power when processing large datasets to deliver better results.

Based on Figure 6, the public has a positive tendency of 49% toward Government Securities (SUN). Several topics that attract public attention in investing in Government Securities include security and risk-free default, as people view SUN as a safe investment because the state guarantees its interest and principal payments and have no default risk; contribution to national development, as the sense of nationalism and contributing to the nation's development is a significant factor in investing in SUN, with people taking pride in purchasing SUN; passive income per period, as periodic coupon payments reassure the public that their investments are safe and managed by the government, seen as a stable passive income source; and ease of accessibility, as with 26 distribution partners available, people can easily access SUN purchases, and additionally, SBR offers an early redemption service.

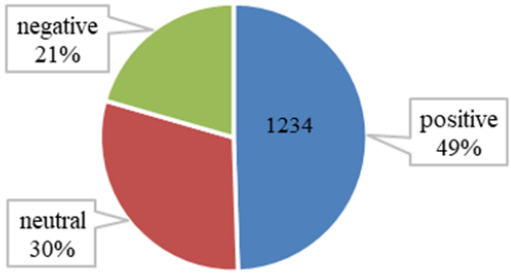


Figure 6. Percentage and Number of Data for Each Sentiment Category

Neutral sentiment accounts for 30% of the entire dataset, showing that some tweets provide simple information about SUN issuance and general explanations. Topics that attract the public's attention in investing in SUN include information on SUN issuance, as distribution partners frequently share tweets regarding the issuance of various SUN series to the public; complaints and services, as distribution partners often assist the public in resolving issues related to SUN transactions, responding directly or guiding them through alternative procedures; and promotion of SUN series advantages, as distribution partners provide information on the characteristics of SUN series being traded, along with issuance details.

The smallest negative sentiment category indicates that the public generally views SUN as a good investment instrument. However, these concerns should still be considered to avoid significant worries and to boost public participation in purchasing SUN. Topics that attract the public's attention in investing in SUN include relatively low return, as people view the return on SUN investments as lower than corporate bonds, with the popularity of mutual funds and other digital investments offering more attractive returns; management mechanism of Tapera fund, as the large proportion of SUN in BP Tapera's financial asset portfolio makes the public view Tapera's fund as a way to finance government programs and the state budget deficit; information and literacy, as public confusion regarding the technical details of SUN persists, despite the abundance of information provided and the increasing financial literacy in Indonesia; and integrity and digital track record, as increasing corruption cases make the public doubt the integrity of government officials, and although concerns about principal and coupon payments on SUN are low, concerns about state debt remain, with President Jokowi's digital track record, particularly regarding bonds in 2012, drawing public attention.

3.6. Strategy Recommendations

Several key points need to be considered based on the analysis that has been conducted. The sale of Government Bonds (SUN), held bi-weekly, has shown stable results below. This strategy has effectively maintained the domestic debt market and supported the economic condition. To date, the issuance of Government Securities (SBN) has proven to be an effective financing instrument to close the budget deficit and support the State Budget (APBN). Below is the issuance of SUN by quarter from 2017 to 2024. Indonesia's conventional financial literacy stands at 65.43%, higher than its Islamic financial literacy at 39.11%, despite the significant potential of the Islamic capital market [19].

In 2022, 77.02% of Indonesians were connected to the internet, with 89.15% of users engaging in activities through social media services [20]. The large opportunity from social media users can be utilized to spread marketing strategies with Key Opinion Leaders (KOLs). Examples of KOLs in the field of finance and investment are Vina Muliana, Felicia, and Chornella [21]. The public's negative sentiment regarding SUN, consisting of 70 tweets or about 13.64%, relates to the Tapera policy as one of the government strategies for accumulating funds through SUN. The government can provide transparent explanations to the public to reduce the negative stigma related to allocating Tapera funds with strategies like the one mentioned above. The investment portfolio proportions should also not be fixed and should be adjusted to the current economic conditions, allowing for the addition of deposit proportions when the BI rate rises, etc. Although the government will not be able to satisfy all parties, optimizing the explanation of the 3% salary fund and the large proportion of SUN as accountability will help reduce public skepticism.

4. CONCLUSION

The conclusions drawn from this study are as follows. First, the BERT model, specifically IndoBERT, demonstrated higher accuracy than the lexicon-based method using InSet across most datasets. Based on the entire tweet dataset, the BERT model achieved the best performance with an accuracy of 70.28%, while the lexicon-based method with InSet achieved 55.77%. Similarly,

in the yearly tweet datasets, the BERT model consistently outperformed the lexicon-based method, except in 2021. Second, The lexicon-based method provides more consistent results, achieving stable accuracy across various dataset sizes while requiring lower computational resources. In contrast, the BERT model shows varying performance, excelling with complex datasets but necessitating higher computational power to process large amounts of data to produce better results. Third, public sentiment toward Government Securities (SUN) on the social media platform X tends to be 49% positive, 30% neutral, and 21% negative. Based on existing opportunities, the government can utilize social media with Key Opinion Leaders (KOLs) and transparency in explaining the Tapera policy. This will maximize public participation in SUN investment in Indonesia. Some recommendations for future consideration are as follows. First, researchers may combine both methods for research involving large datasets: using Lexicon-based InSet labeling as the ground truth for training data and performing sentiment analysis with the BERT model. Second, future researchers may consider upgrading to Google Colab Pro or Colab Pro+ to optimize machine learning model training, as these options provide enhanced computational resources.

ACKNOWLEDGEMENTS

The author expresses gratitude to Allah SWT and thanks to the professors and staff of the Department of Actuarial ITS, especially to Mrs. Ulil Azmi and Mrs. Nia, my advisors, for their invaluable assistance. I also thank my parents, family, friends, and all those who have supported and prayed for me throughout this journey.

REFERENCES

[1] L. L. Sari and M. Marselina, "Efektivitas dan Kontribusi Surat Berharga Negara terhadap Pembiayaan Defisit Anggaran Pendapatan dan Belanja Negara di Indonesia Tahun 2012-2021," *Journal on Education*, vol. 6, no. 1, pp. 10 683–10 694, Dec. 28, 2023.

[2] S. Thomas, Y. Yuliana, and Noviyanti P, "Study Analisis Metode Analisis Sentimen pada YouTube," *Journal of Information Technology*, vol. 1, no. 1, pp. 1–7, 2021. DOI: 10.46229/jifotech.v1i1.201.

[3] D. Musfiroh *et al.*, "Analisis Sentimen terhadap Perkuliahan Daring di Indonesia dari Twitter Dataset Menggunakan InSet Lexicon: Sentiment Analysis of Online Lectures in Indonesia from Twitter Dataset Using InSet Lexicon," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 1, no. 1, pp. 24–33, Mar. 6, 2021. DOI: 10.57152/malcom.v1i1.20.

[4] A. Setiawan, "Analisis Sentimen Masyarakat di Twitter terhadap Kejadian Bom Bunuh Diri Polsek Astana Anyar Menggunakan Algoritma SVM dengan Leksikon Vader dan Inset," BA thesis, Fakultas Sains dan Teknologi UIN Syarif HIdayatullah Jakarta, Jan. 19, 2024.

[5] A. Faizal, A. S. Y. Irawan, and D. Juardi, "Perbandingan Lexicon Based dan Naïve Bayes Classifier pada Analisis Sentimen Pengguna Twitter terhadap Gempa Turki," *INTECOMS: Journal of Information Technology and Computer Science*, vol. 6, no. 2, pp. 1037–1048, Nov. 21, 2023. DOI: 10.31539/intecoms.v6i2.7360.

[6] Z. A. Sriyanti, D. S. Y. Kartika, and A. R. E. Najaf, "Implementasi Model Bert pada Analisis Sentimen Pengguna Twitter terhadap Aksi Boikot Produk Israel," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 12, no. 3, pp. 2335–2342, Aug. 3, 2024. DOI: 10.23960/jitet.v12i3.4743.

[7] C. Sahoo, M. Wankhade, and B. K. Singh, "Sentiment analysis using deep learning techniques: A comprehensive review," *International Journal of Multimedia Information Retrieval*, vol. 12, no. 2, p. 41, Dec. 2023. DOI: 10.1007/s13735-023-00308-2.

[8] M. F. Ramadhan and B. Siswoyo, "Mengenal Model BERT dan Implementasinya untuk Analisis Sentimen Ulasan Game," *Prosiding SISFOTEK*, vol. 8, no. 1, pp. 395–398, Oct. 27, 2024.

[9] R. M. R. W. P. K. Atmaja and W. Yustanti, "Analisis Sentimen Customer Review Aplikasi Ruang Guru Dengan Metode BERT (Bidirectional Encoder Representations from Transformers)," *Journal of Emerging Information System and Business Intelligence (JEISBI)*, vol. 2, no. 3, pp. 55–62, Jul. 12, 2021.

[10] B. Kurniawan, A. A. Aldino, and A. R. Isnain, "Sentimen Analisis Terhadap Kebijakan Penyelenggara Sistem Elektronik (PSE) Menggunakan Algoritma Bidirectional Encoder Representations From Transformers (BERT)," *Jurnal Teknologi dan Sistem Informasi*, vol. 3, no. 4, pp. 98–106, 2022. DOI: 10.33365/jtsi.v3i4.2204.

[11] S. Tabinda Kokab, S. Asghar, and S. Naz, "Transformer-based Deep Learning Models for the Sentiment Analysis of Social Media Data," *Array*, vol. 14, p. 100 157, Jul. 2022. DOI: 10.1016/j.array.2022.100157.

[12] V. Chandradev, I. M. A. D. Suarjaya, and I. P. A. Bayupati, "Analisis Sentimen Review Hotel Menggunakan Metode Deep Learning BERT: Analisis Sentimen Review Hotel menggunakan Metode Deep Learning BERT," *Jurnal Buana Informatika*, vol. 14, no. 02, pp. 107–116, Oct. 1, 2023. DOI: 10.24002/jbi.v14i02.7244.

[13] A. Ardiansyah *et al.*, "Analisis Sentimen terhadap Pelayanan Kesehatan Berdasarkan Ulasan Google Maps Menggunakan BERT," *Jurnal FASILKOM*, vol. 13, no. 2, pp. 326–333, Aug. 31, 2023. DOI: 10.37859/jf.v13i02.5170.

[14] M. Amien and G. F. Gunawan, "BERT dan Bahasa Indonesia: Studi tentang Efektivitas Model NLP Berbasis Transformer," *ELANG: Journal of Interdisciplinary Research*, vol. 1, no. 2, pp. 132–140, Jan. 30, 2024. DOI: 10.32664/elang.v1i02.

[15] S. Sutoyo, "Pengaruh Surat Berharga Negara (SBN) terhadap Pertumbuhan Ekonomi Indonesia," *Lentera : Jurnal Ilmiah Sains, Teknologi, Ekonomi, Sosial, dan Budaya*, vol. 6, no. 2, pp. 86–89, May 31, 2022.

[16] B. Juanda and S. Gladiola, "Analisis Keberlanjutan serta Pengaruh Surat Berharga Negara dan Faktor Lainnya terhadap Pertumbuhan Ekonomi di Indonesia," *Indonesian Treasury Review: Jurnal Perbendaharaan, Keuangan Negara dan Kebijakan Publik*, vol. 7, no. 3, pp. 239–254, Sep. 30, 2022. DOI: 10.33105/itrev.v7i3.529.

[17] S. Arifin and H. Adjie, "Keamanan Surat Utang Negara yang Dijual Secara Lelang Ditinjau dari Asas Akuntabilitas," *E-Jurnal SPIRIT PRO PATRIA*, vol. 8, no. 2, pp. 82–88, Sep. 30, 2022. DOI: 10.29138/spirit.v8i2.2151.

[18] C.-R. Ko and H.-T. Chang, "LSTM-based Sentiment Analysis for Stock Price Forecast," *PeerJ Computer Science*, vol. 7, e408, Mar. 11, 2021. DOI: 10.7717/peerj-cs.408.

[19] K. F. Ariani, T. I. Rahmawati, and D. V. Anggraini, "Peningkatan Literasi Keuangan Masyarakat Pedesaan Guna Mendorong Tingkat Inklusi Keuangan Indonesia Perspektif Hukum Perbankan," *Jurnal Multidisiplin Ilmu Akademik*, vol. 1, no. 6, pp. 118–128, 2024. DOI: 10.61722/jmia.v1i6.2874.

[20] A. Hermawansyah and A. R. Pratama, "Analisis Profil dan Karakteristik Pengguna Media Sosial di Indonesia Dengan Metode EFA dan MCA," *Techno.Com*, vol. 20, no. 1, pp. 69–82, Feb. 9, 2021. DOI: 10.33633/tc.v20i1.4289.

[21] D. A. Editya. "Tinjauan atas Efektivitas Penggunaan Key Opinion Leader (KOL) dalam Penjualan Surat Utang Negara Ritel seri SBR011. "[Online]. Available: https://arxiv.org/abs/2208.12619, pre-published.